

Proof of SVD

Key Fact: $N(A^T A) = N(A)$ (recitation)

Let A be $m \times n$ and $r = \text{rank}(A)$

Recall that $A^T A$ is symmetric and positive semi-definite.

Thus its eigenvalues are real and nonnegative.

Denote them by $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$.

How many are 0? # 0 eigenvalues = $\dim N(A^T A)$

Why? Since $A^T A$ is diagonalizable (as it is symmetric),

$$AM(\lambda=0) = GM(\lambda=0) = \dim E_0(A^T A) = \dim N(A^T A)$$

(AM = algebraic multiplicity & GM = geometric multiplicity)

Since A is $m \times n$, we have $\dim N(A^T A) = \dim N(A) = n - r$

Hence $\lambda_1 \geq \dots \geq \lambda_r > 0$ and $\lambda_{r+1}, \dots, \lambda_n = 0$.

Let $\sigma_i = \sqrt{\lambda_i}$ $\forall 1 \leq i \leq r$ (singular values)

Let $\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \\ & & & 0 \end{bmatrix}$ be $m \times n$

Now let's build V .

By the Spectral Theorem, $A^T A$ is orthogonally diagonalizable:

$$A^T A = Q \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} Q^T$$

where Q is orthogonal whose columns are orthonormal

eigenvectors $v_1, \dots, v_n$ of $A^T A$ corresponding to $\lambda_1, \dots, \lambda_n$

Let $V = Q$.

Now let's build U .

Strategy:

- Construct U out of A, Σ , and V
- Show that the columns of U satisfy the SVD properties (i.e. they are evecs of AA^T and orthonormal)

Let $U = [u_1 \dots u_m]$, where

Let $u_i = \frac{Av_i}{\sigma_i}$ for $1 \leq i \leq r$ (first r columns of U)

and $\{u_{r+1}, \dots, u_m\}$ is an orthonormal basis of $N(A^T)$

Lemma:

- u_i is an eigenvector for AA^T corresponding to λ_i
- U is orthogonal

Proof:

a). For $1 \leq i \leq r$,

$$AA^T u_i = AA^T \left(\frac{Av_i}{\sigma_i} \right) = \frac{1}{\sigma_i} A(A^T Av_i) = \frac{1}{\sigma_i} A(\lambda_i v_i) = \lambda_i \left(\frac{Av_i}{\sigma_i} \right) = \lambda_i u_i$$

$\Rightarrow u_i$ is evector of AA^T corresp to λ_i

• For $r+1 \leq i \leq m$, $AA^T u_i = A(A^T u_i) = A(0) = 0 = 0 \cdot u_i$

$\Rightarrow u_i$ is evector of AA^T corresp to $0 = \lambda_i$

$\uparrow$
 $u_i \in N(A^T)$

b) We first show $\{u_1, \dots, u_r\}$ is orthonormal

$$\begin{aligned} u_i \cdot u_j &= \left(\frac{Av_i}{\sigma_i} \right) \cdot \left(\frac{Av_j}{\sigma_j} \right) = \frac{1}{\sigma_i \sigma_j} (Av_i)^T (Av_j) = \frac{1}{\sigma_i \sigma_j} v_i^T (A^T A v_j) \\ &= \frac{1}{\sigma_i \sigma_j} v_i^T (\lambda_j v_j) = \frac{\sqrt{\lambda_j}}{\sqrt{\lambda_i \lambda_j}} v_i \cdot v_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \end{aligned}$$

Since $\{v_1, \dots, v_n\}$ is orthonormal

Next, $\{u_{r+1}, \dots, u_m\}$ is orthonormal by def'n.

It remains to show that $u_i \cdot u_j = 0 \ \forall \ 1 \leq i \leq r, \ r+1 \leq j \leq m$.

Let $1 \leq i \leq r$ and $r+1 \leq j \leq m$.

Notice, $u_i = A \left(\frac{v_i}{\sigma_i} \right) \in C(A)$ and $u_j \in N(A^T)$ by definition

Since $N(A^T) \perp C(A)$, we have $u_i \cdot u_j = 0$ 

So we've built matrices U, Σ, V out of eigenvalues and eigenvectors. We now show that $U \Sigma V^T = A$

Theorem: Let U, Σ , and V be as above. Then $A = U \Sigma V^T$

Moreover, any SVD of A can be obtained in this way.

Proof: Since $V^T V = I$, multiplying $A = U \Sigma V^T$ by V (on the right) gives $AV = U \Sigma$.

So it suffices to verify $AV = U \Sigma$

$$\begin{aligned}
 U\Sigma &= [u_1 \dots u_m] \begin{bmatrix} \sqrt{\lambda_1} & \dots & 0 \\ 0 & \sqrt{\lambda_r} & \dots & 0 \end{bmatrix} \\
 &= [\sqrt{\lambda_1} u_1 \dots \sqrt{\lambda_r} u_r \ 0 \dots 0] \\
 &= \left[\sqrt{\lambda_1} \frac{Av_1}{\sqrt{\lambda_1}} \dots \sqrt{\lambda_r} \frac{Av_r}{\sqrt{\lambda_r}} \ 0 \dots 0 \right] \\
 &= [Av_1 \dots Av_r \ 0 \dots 0]
 \end{aligned}$$

Here note that since $\lambda_i = 0 \ \forall \ r+1 \leq i \leq n$, $(A^T A)v_i = \lambda_i v_i = 0 \ \forall \ r+1 \leq i \leq n$. Hence $v_i \in N(A^T A)$. By the Key Fact, $v_i \in N(A)$ and so $Av_i = 0 \ \forall \ r+1 \leq i \leq n$. Hence:

$$\begin{aligned}
 &= [Av_1 \dots Av_r \ Av_{r+1} \dots Av_n] \\
 &= AV
 \end{aligned}$$

Hence $AV = \Sigma U \Rightarrow A = U\Sigma V^T$

We now show that any SVD of A is built out of evals of $A^T A$ and evecs of $A^T A$ and AA^T .

More precisely, if $A = U\Sigma V^T$ is an SVD, then:

- σ_i^2 is eval of $A^T A \ \forall i$
- columns of V are eigenvectors of $A^T A$
- columns of U are eigenvectors of AA^T

First note:

$$A^T A = (U \Sigma V^T)^T (U \Sigma V^T) = V \underbrace{\Sigma^T U^T U}_{=I} \Sigma V^T = V \Sigma^T \Sigma V^T$$

$$\text{and } \Sigma^T \Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix} \cdot \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_r^2 & & 0 \\ & & & \ddots & \\ & & & & 0 \end{bmatrix}$$

Since $\Sigma^T \Sigma$ is diagonal, by the diagonalization theorem, $\sigma_1^2, \dots, \sigma_r^2$, (and possibly 0) are the eigenvalues of $A^T A$ and the columns of V are the corresponding eigenvectors

$$\text{Similarly, } A A^T = (U \Sigma V^T)(U \Sigma V^T)^T = U \Sigma V^T V \Sigma^T U^T = U \Sigma \Sigma^T U^T$$

$\Rightarrow$ columns of U are eigenvectors of $A A^T$ corresp to $\sigma_1^2, \dots, \sigma_r^2$, (and possibly 0)



Finally note that we have implicitly shown that:

- $\{v_1, \dots, v_r\}$ is an orthonormal basis of $C(A^T)$
- $\{v_{r+1}, \dots, v_n\}$ is an orthonormal basis of $N(A)$
- $\{u_1, \dots, u_r\}$ is an orthonormal basis of $C(A)$
- $\{u_{r+1}, \dots, u_m\}$ is an orthonormal basis of $N(A^T)$